

EVALUATION AND GRADING OF PHILOSOPHY ESSAYS AT THE INTERNATIONAL PHILOSOPHY OLYMPIAD

Jonas Pfister
Gymnasium Neufeld Bern
pfister.jonas@gmail.com

Received: 11 July 2020

Accepted: 23 December 2020

Abstract

The correct evaluation of philosophy essays is an important topic in the didactics of philosophy. Two central questions are which evaluation criteria should be used and how they should be applied to reach fair grades. Focusing in this article on the essay competition of the International Philosophy Olympiad (IPO), the list of five evaluation criteria for an argumentative essay at the IPO is discussed. It is argued that these criteria should be specified, their status should be clarified, and a grading method should be developed. The article contains an appendix with a very short guide for evaluators at the International Philosophy Olympiad.

Keywords: Evaluation Criteria, Fair Grades, Essay, International Philosophy Olympiad

The correct evaluation of philosophy essays is an important topic in the didactics of philosophy. It is important because it involves questions about concepts that touch the fundamentals of what philosophy is, epistemological questions about recognition of good philosophy, and ethical questions of fairness in grading. I focus here on the essay competition of the International Philosophy Olympiad (IPO), an essay competition for high school students that has been held since 1993. I first explain the concept of a philosophical essay and the criteria of evaluation at the IPO. I discuss some of the problems with these criteria and propose some ideas on how to improve them. Then I discuss some problems in the grading process and propose some solutions. I argue that a grading method should be developed to increase the validity and reliability of the grades.¹

1. The concept of a philosophy essay and evaluation criteria at the IPO

What is a philosophical essay? There exist different concepts of philosophical essays. One can distinguish between a more literary form in the tradition of Montaigne and a more argumentative form in the tradition of Francis Bacon. Given that one of the criteria of evaluation at the IPO is about argumentation, it is this second, argumentative form that is intended in the

¹ I would like to thank an anonymous reviewer for very helpful constructive criticism. It helped me see the flaws in the structure and the argumentation in a previous version of this article. That first version did not satisfy the criteria of a good philosophical essay by any standards. The constructive criticism led to a reorganization of the whole essay and to a development of my arguments. I hope that now at least the general line of argumentation is fairly clear.



competition and that is the focus here.

As a general characterization of an argumentative philosophy essay, we may take the one by Jay F. Rosenberg in his introductory book *The Practice of Philosophy*:

A philosophical essay is neither a research paper [...], nor is it a literary exercise in self-expression. It does not deal with feelings or impressions. It is not a report or summary. Fundamentally, it is the *reasoned defense of a thesis*. That is, there must be some point or points to be *established* in the essay, and considerations must be offered in *support* of them in such a way that the considerations can be seen to support them. (Rosenberg 1996: 56)

Similarly, in his guide on how to write a philosophy essay for IPO contestants, Frank Murphy writes that “the critical components of any philosophical inquiry have always been to craft a thesis [...] and persuade a listener or reader to accept one’s thesis through honest, logical, and thorough argumentation” (Murphy 2017: 53).

Given this characterization, we can say that a good philosophical essay is one that, at least, *argues well for a thesis*. The demands for such philosophical writing include, as Rosenberg mentions, “clarity of exposition, precision of statement, organization of ideas, and logical rigor and consistency” (Rosenberg 1996: 56). We thus have an understanding of the argumentative essay as it is widely used in teaching philosophy and at the IPO. I therefore do not share Marc Foglia’s skepticism about the concept (Foglia 2020, this issue). However, I share some of his concerns about the evaluation criteria at the IPO.

What are the evaluation criteria of the essays at the IPO? They are the following five: relevance to the topic, coherence, philosophical understanding of the topic, persuasive power of argumentation, and originality.² These criteria are tailored to the specifics of the competition which includes the writing of an essay in four hours on a chosen topic out of four (philosophical quotations or questions) without the use of any material except for a mono- or bilingual dictionary.

These five criteria have a great record of success. Up to this year, they have been used in twenty-eight editions of the IPO and have, as far as I know, stayed the same during all these years. This is an impressive record, considering also the varieties of philosophical cultures and traditions that come together at the competition. However, there are also some problems with these criteria, some of which I will now discuss.

2. Problems with the evaluation criteria

I would like to mention three types of problems for the evaluation criteria of the IPO. First, problems concerning their justification: Are all of the five criteria justified? Second, problems concerning their meaning: What do the criteria mean exactly? Can they be interpreted differently, and are all of the interpretations to be included? Third, problems concerning their application: How are the criteria applied? Do all of them have the same status? What is their relative weight?

² See the statutes at: <http://www.philosophy-olympiad.org/>

(1) Justification of the criteria: No justification for the criteria is explicitly given in the statute of the IPO. However, several of them are implicitly justified: Given what a good argumentative philosophy essay is, we can say that it is justified to use the criteria of relevance to the topic, coherence, argumentation, and philosophical understanding.³ But, as I will argue now, one aspect of the criterion about argumentation and the criterion of originality are not justified.

The criterion about argumentation demands “persuasive” power. But to persuade is not a necessary condition for a good argument. An argument may be excellent according to objective standards – deductively valid and consisting of premises that are true or that may be reasonably held – and yet one might not be persuaded by the argument because one does not believe that one of the premises which may reasonably be held is in fact true. I may acknowledge that the argument is excellent according to these standards, while at the same time not believing that all of the premises are true. Therefore, persuasiveness is not a necessary condition for a good argument. Furthermore, persuasiveness is not a genuinely philosophical aim, and it even may be anti-philosophical. A bad argument – an argument that is invalid or unsound – may be very persuasive, using rhetorical techniques of pleasing the reader, for example. “Power of argumentation” should therefore not be meant to be “persuasive” power.

The criterion of originality is problematic in the context of a high school competition. First, what is original is radically dependent on background and context. For a high school student not having enjoyed a course in epistemology, developing a radical skeptic, empiricist or rationalistic view about knowledge can be very original whereas for another student it might just be the reproduction of course material. The educational and philosophical background of the students is not generally known to the assessors which makes it in principle difficult if not impossible in some cases to apply the criterion. Second, what someone takes to be original in a philosophy essay is heavily dependent on subjective evaluations. It is therefore a highly subjective criterion. And a highly subjective criterion should not be used in grading a philosophy paper. Such an argument has been put forward by Frosina Postoloska in discussions at the IPO.

(2) Meaning of the criteria: The meaning of each of the five criteria is not clear. For example, does relevance to the topic relate to the whole essay or to its parts or to both? In what way has the essay to be original? Must it design a new solution to a problem? (Probably not.) What counts as philosophical understanding? Must the essay refer to positions in the history of philosophy? In case the topic is a quote, must the essay give a correct interpretation of it? (It is generally taken at the IPO not to include this, but it is not stated anywhere.) Is the criterion of coherence to be understood to apply to the exposition of the essay, the use of terms, or the claims made? What does the criterion of the power of argumentation include? (At least, it should not include persuasiveness, see above.) All of the criteria in their present, general form

³ I am concerned here about whether or not it is justified to have these criteria, not about the more fundamental question about what such justification could be. I see at least three ways in which a criterion can lack justification: a) It is not adequately related to what a good argumentative philosophy essay is. b) It is not something that high school students can reasonably be taken to try to achieve. c) It cannot be consistently applied in principle. (In the context of education, the criteria of evaluation can be justified by reference to educational goals, that is, in the end, being an independent thinker cooperating in society, see Campbell 1988, Part Two).

need first to be interpreted before they can be applied.

(3) Application of the criteria: More problems arise when one tries to apply the criteria. One concerns the status of the criterion of relevance to the topic. Is it a criterion like any other? No, it cannot be. It should rather be a *prerequisite* for a good essay. This is a point that has been repeatedly made by Barbara Conrad in discussions of the International Jury of the IPO. If one does not treat relevance to the topic as a prerequisite (or ascribes to the criterion a very heavy weight), an essay that has no relevance for the chosen topic whatsoever could nevertheless receive a high grade, and since such an essay could have been prepared in total in advance, such grade would not be justified. Therefore, relevance (in one sense) should be a prerequisite for a good essay.

A second problem concerns the relative weight of the criteria. Do they all have the same weight? According to the rubric of Floris Velema they do (see Foglia 2020: 132). But such equal weighing is not stated in the statute, and it is doubtful whether it would lead to acceptable grades. For example, an essay that shows “insight and competence, as well as a clear understanding of thinkers, concepts, theory, etc.” but voicing no original viewpoint would receive the same grade as an essay showing originality (“character”, being “colorful” and expressing “personality”) but no philosophical understanding. That seems to be the wrong result.

3. Possible solutions to the problems

One way to deal with the problems is to keep the five criteria and to specify their meaning, clarify their status, and determine their weight. This is the more conservative way. Another, more progressive – but not necessarily better – way would be to design a new list of criteria.

Going the more conservative way, keeping the five criteria, one could do the following: a) clarify that relevance to the topic is in one sense a prerequisite for a good essay, b) specify that the power of argumentation does not include persuasiveness but includes validity and soundness of arguments, c) specify that coherence applies both to the exposition of the essay (its structure) and to the use of terms, d) specify that philosophical understanding does not require but may include knowledge of the history of philosophy and, in the case of a quote, does not require a correct interpretation of the quote, e) specify that originality means that the author develops his or her own thoughts rather than repeating course material (even if it may be hard to know whether this is the case or not), and f) leave the relative weight of the criteria open or, better, determine the relative weight. See the appendix for a more detailed possible specification of the criteria within the given system of evaluation with the five criteria (going the conservative way).

The more progressive way would be to design a new list of criteria. Inspiration can be found in the literature. In an early article, James Campbell (1988) mentions four criteria he takes to be present, consciously or not, in our actual grading of philosophy papers: (1) the student’s conformity with our position or the adoption of a sensible conclusion, (2) the student’s correct reproduction of position or demonstration of a clear and critical understanding, (3) the ability of the student to present and defend his or her position, (4) the amount of work rendered (Campbell 1988, Part One: 7-8). Campbell argues that we should use the first criterion as little

as possible and aim at neutrality in grading (Campbell 1988, Part Two: 4). Similarly, Jim Pryor writes the following at the attention of the students in his “Guidelines on Writing a Philosophy Paper” (2012): “We do not judge your paper by whether we agree with its conclusion. In fact, we may not agree amongst ourselves about what the correct conclusion is. But we will have no trouble agreeing about whether you do a good job arguing for your conclusion.” Pryor uses three basic criteria that he formulates at the attention of the students:

You’ll be graded on three basic criteria:

1. How well do you understand the issues you’re writing about?
2. How good are the arguments you offer?
3. Is your writing clear and well-organized?

These three criteria roughly correspond to those from the IPO: philosophical understanding of the topic, power of argumentation, and coherence, understood as a coherent structure of the text. But Pryor’s list is superior to it, first, because it does not include persuasiveness (see Pryor’s comment about agreeing with the conclusion above), second, because the criteria do not overlap, and third, because it additionally includes an important aspect of a philosophical text: clarity.

Can these criteria be justified? Yes, first, by what we consider to be a good philosophy essay (see above). Second, viewing the competition of high school students in the broader context of education, they can also be justified by our understanding of what education is and should be. James Campbell argues that the criterion of being able to correctly reproduce and critically engage with a given position justified because it represents the capacity to “open up to the ideas and perspectives of other individuals” and that the criterion of being able to defend one’s position is justified because it represents the intellectual capacities to solve problems. Both of these capacities are part of the view of education as preparing human beings for “an independent yet cooperative adulthood” (Campbell 1988, Part Two: 3-4). I would argue that the third criterion is justified because it represents the intellectual capacity of clear thinking, and this is part of an independent adult life as well.

If this list of only three criteria (or any other list) is to be used, all of the criteria need to be specified so that they can be applied consistently. The application of the criteria to reach grades raises further challenges.

4. Grading

How should the criteria of evaluation be applied to reach grades? To answer this question, I start with some remarks about the application of criteria in the evaluation of student achievement in general. Before applying criteria of evaluation in order to reach and assign a grade to a student, one should have, according to Thomas Haladyna (1999: ix):

1. an idea about what a grade means,
2. an understanding of the purposes of grading,
3. a set of personal beliefs and proven principles that we will use in teaching and grading,

4. a set of criteria on which the grade is based, and, finally,
5. a grading method, which is a set of procedures that we consistently follow in arriving at each student's grade.

At the IPO, the grading process is divided into three stages: the first and second stages are carried out by the International Jury, the third stage by the Steering Board who is in charge of the distribution of prizes and medals. Grades are given from 1 to 10 using intervals of .5. They are given the following meaning:⁴

- 7.5-10 points means: I suggest this essay for the next stage.
- 6-7 points means: I myself don't suggest this essay, but I will agree if somebody else selects this essay.
- 1-5 points means: I suggest that this essay should not be accepted for the next stage.

There are at least two problems with these meanings. First, a minor problem of phrasing. The formulation "I will agree if somebody else selects this essay" is confused. If only one person out of four selects the essay and the others do not, there is no reason to agree with suggesting this essay for the next stage. And if the average of the grades is above a certain level – 7 in the first stage – I need not agree *in the matter* with the judgment that this particular essay should reach the next stage; I will agree that it should reach the next stage *because of the rule that an essay having an average grade of 7 reaches the second stage*. And whether or not I personally agree with this rule is irrelevant. This minor problem of formulation can easily be solved by deleting the phrase.

Second, a problem of greater importance. The transitions from the first to the second stage and from the second to the third stage are remarkably different. Going on to the second stage only means that the essay receives additional gradings. Going on the third stage means that the essay will be read by the Steering board and be taken into consideration for a medal. These are two different things. I might for example come to the judgement that an essay should receive additional readings without believing that it should be a candidate for a medal. What grade should I then give to the essay in the first round? Additionally, for some essays, it can be quite difficult and cumbersome to reach a precise grade, for example to be able to clearly decide between 7.5 or 8. Requiring at the first stage an exact grade diminishes the efficiency (or the reliability) of the general grading process.

A possible solution to this second problem is to have separate procedures for the two transitions. In the first transition, the purpose is to find out which essays should get additional readings. This can be done with a simple "triage" method using three grades: 1 for pass, 2 for undecided, 3 for not pass. One would then have to install a rule for how many "passing" individual grades the essay would need to go on to the second stage. Such a triage method can be used for grading in general, as suggested by William Rapaport: for each item or criterion, full credit is given if and only if it is substantially correct, minimal credit if and only if it is substantially incorrect, and partial credit if it is neither (Rapaport 2011: 347).

⁴ See the statutes at: <http://www.philosophy-olympiad.org/>

In the second transition, from the second to the third stage, the purpose is to find the candidates for medals. For this, *all* of the essays of the second stage should be read again. And it is only after this second reading that the precise grades should be given.

What are the purposes of the evaluation and grading of the essays at the IPO? The primary purpose is to determine who receives awards. This means that the evaluation is clearly *summative*, expressing a judgment of the achievement, as opposed to a *formative* evaluation, giving feedback to the students in order to improve their achievement (see Scriven 1967). The handing out of an award communicates the achievement to the student, to the community of the IPO, to the local school, to regional and national communities in the home country of the student, and to the universities and employers that may accept the young laureate in the future.

What principles do we follow in grading philosophy papers? The arguably most basic principle in grading student achievement is this: grades should be fair! As a minimal requirement for a fair grade for an achievement by a student, one can say that a) each student was able to show his or her capacities, b) the grade reflects the student's capacities, c) the grade is arrived at through the application of objective criteria, d) these criteria are known to all students, e) these same criteria are used for the evaluation of the achievements of all students, and f) each of the students receives the same tasks or at least tasks of the same difficulty (Pfister 2014: 81; see also Haladyna 2019: 6; Weis 1995; Close 2009; McCrickerd 2012; Burkholder 2015).

There should be a list of criteria, and these criteria should be as objective as possible. That means, in the case of the criteria of the IPO, that these need to be specified (see above). The criteria and their specification should be communicated to the participants.

Criteria by themselves do not state how they are to be applied. For this, a grading method needs to be designed, that means a set of procedures to be consistently followed to arrive at the students' grades. The central concepts in the grading of student achievement are *validity* and *reliability*, where validity is the accuracy of a grade's reflection of student achievement and reliability is the degree of random error that might affect validity (Haladyna 2019: 1). The grading method should thus aspire to valid and reliable grades.

The more objective the criteria are, the more valid and reliable the grades will be. A subjective element will always remain, be it only because the essay has to be evaluated as a whole text, as one single product of a creative human mind. It would therefore be wrong to exclude subjective elements altogether. But the criteria should be as objective as possible.

Striving for validity also means that the relative weight of the criteria should be determined. This can be done by the simple indication of a percentage, the weight of all criteria adding up to one hundred percent.

Once this is done, one could design a rating grid and rubric. Inspiration can be found in Linda Farmer's list: clear thesis, thesis support, accurate exegesis, critical reasoning, objections & replies, original contribution (see Farmer 2003) and Maralee Harrell's list: argument (thesis, premises, support, counterarguments), understanding, analysis, synthesis, and creation (see Harrell 2005).

Reliability can be augmented by minimizing cognitive biases. Among the known biases are the primacy-effect (the first impression is carried over to the rest), the halo-effect (the

personality traits of the author are carried over to the essay), the effect of eating and breaks (we tend to be more lenient/less severe after a break). Furthermore, evaluators may be too severe (underrating) or too lenient (overrating) or tend to compression (giving all the essays the same grade). If we want the grades to be reliable, the effects of such biases need to be minimized. Being aware of them can already help, as well as small changes such as reading the essays in a different order and taking breaks regularly.

5. Reply to some objections

Against the proposed specifications of the criteria (or a possible new list of precise criteria) and the design of a grading method to reach valid and reliable grades, one might have some objections. One of the main objections that one might have is that such precise criteria and such a grading method would lead the students to a mechanical way of writing the essay, and the contest would lose what is precisely so special and worth preserving about it, namely that the essay is an expression of the author's thoughts unconstrained by formal requirements. One might fear that the specified criteria would constrain the thinking process of the students and that their application would lead to a uniformity in the products, thus leading to essays of lower quality and much less fun, both for the students and the evaluators.

This objection is in part based on wrong premises. The specification of criteria of evaluation cannot constrain the expression of one's thoughts. The freedom to express any thought whatsoever remains even if one would require a very specific form of its expression. Following certain rules does not impair one's thinking but rather supports it. The objection, it seems to me, also underestimates the potential of creation inherent to language. The specification of the criteria would not preclude, but rather encourage the designing of new examples, cases, and thought experiments. It is therefore unlikely that the specification of the criteria would lead to less interesting essays. However, a certain uniformity in the structuring of the essays would probably be achieved – but that would be very welcome for it would increase the quality of the essays.

Another objection that one might have is that the system of the criteria and their specifications could become too complex for the evaluators to handle, and the disagreement among the evaluators could still remain, thus not improving the process of grading.⁵

The danger mentioned by this objection is real. It is possible that a list of precise criteria gets so complex that it requires too many aspects for the evaluators to consider at the same time. This is not only a problem because of cognitive effort and time, but also because the important aspects of an essay – in particular the clear development of an argument – might get submersed under a multitude of minor issues. But the danger can be avoided by designing a list of basic criteria and specifications that can be handled by the evaluators. A list of for example three criteria and three specifications for each in a rubric would seem to me to be a reasonable proposal. Of course, there still could be disagreement among different evaluators about the meaning of these criteria. But it would concern a narrower meaning. And of course, there still will be disagreements among different evaluators about the correct application of these criteria in the particular case. But it would definitely be an improvement compared to the present list

⁵ I would like to thank an anonymous reviewer for raising this objection.

because the range of applications is limited by the specification of the criteria.

I hope to have shown that the criteria of evaluation of the IPO need to be improved in three respects: they need to be specified, their status has to be clarified, and a grading method needs to be developed. If the grades of the essays in the competition are to have a high degree of validity and reliability, then this is a necessary step to take.

References

- Burkholder, Leslie (2015), "Impartial Grading Revisited", *Teaching Philosophy* 38(3), 261–272.
- Campbell, James (1988), "Grading Philosophy Papers", *American Association of Philosophy Teachers News*, Part One, March 1988, 5–8; Part Two, June 1988, 3–6.
- Close, Daryl (2009), "Fair Grades", *Teaching Philosophy* 32(4), 361–398.
- Farmer, Linda L. (2003), "Grading Argumentative Essays", *Teaching Philosophy* 26(2): 125–130.
- Foglia, Marc (2020), "What is an essay? Thoughts on the essay as the form of philosophical expression at the International Philosophy Olympiad (IPO)," *Journal of Didactics of Philosophy* 4(3), 130–136. URL: www.philosophie.ch/jdph
- Haladyna, Thomas M. (1999), *A Complete Guide to Student Grading*, Boston: Allyn & Bacon.
- Haladyna, Thomas M. (2019), "Assigning a Valid and Reliable Grade in a Course", *IDEA paper* 79. URL: https://www.ideaedu.org/idea_papers/assigning-a-valid-and-reliable-grade-in-a-course/
- Harrell, Maralee (2005), "Grading According to a Rubric," *Teaching Philosophy* 28(1), 3–14
- McCrickerd, Jennifer (2012), "What Can Be Fairly Factored into Final Grades?" *Teaching Philosophy* 35(3), 275–291.
- Murphy, Frank (2017), "International Philosophy Olympiad: A Writing Challenge for Young Philosophers. With an Appendix: How to Write a Philosophy Essay. A Guide for IPO Contestants," *Journal of Didactics of Philosophy* 1, 49–66. URL: www.philosophie.ch/jdph
- Pfister, Jonas (2014), *Fachdidaktik Philosophie*. 2. Auflage, Bern/Stuttgart: Haupt/UTB.
- Pryor, James (2012), "Guidelines on Writing a Philosophy Paper," (originally 2003, updated 2012). URL: <http://www.jimpryor.net/teaching/guidelines/writing.html>
- Rapaport, William, J. (2011), "A Triage Theory of Grading: The Good, the Bad, and the Middling," *Teaching Philosophy* 34(4): 347–372.
- Rosenberg, Jay F. (1996), *The Practice of Philosophy. A Handbook for Beginners*. Third edition, Prentice-Hall. (1st ed. 1976)
- Scriven, Michael (1967), "The Methodology of Evaluation," in: R. Tyler, R. Gagné, M. Scriven (ed.), *Perspectives on Curriculum Evaluation*, Chicago: Rand McNally.
- Weis, Gregory F. (1995), "Grading," *Teaching Philosophy* 18(1), 3–13.

Appendix

How to Evaluate a Philosophy Essay at the International Philosophy Olympiad A Very Short Guide for Evaluators

This is a very short guide for evaluators (delegation leaders and teachers) who are members of the International Jury at the International Philosophy Olympiad (IPO) and especially for those who are members for the first time. This guide is not part of the official regulations and it does not state rules. First, it explains some of the fundamentals of the competition and the evaluation procedure. Second, it elucidates the meaning of the five criteria of evaluation. Third, it ends with some remarks about discussing essays with other evaluators.

1. Fundamentals of the competition and evaluation procedure

The IPO has among its objectives to promote philosophical education at secondary level around the world and to increase the interest of high school students in philosophy. The awarding of medals and honorable mentions to the best contestants of the competition serves these objectives. It gives a motivating feedback to the winners and it helps to promote philosophy education in the countries to which the winners return.

As part of the International Jury, you will be asked to assess a number of essays in order to help determine who will be awarded a medal or honorable mention. You will only read a very small number of all the essays in the competition. The vast majority of the essays you will not get to read. And this leads to an important point to keep in mind: The essays you will not read might be of lower or of better quality than the essays you will read, and you will not know whether the first or the second applies.

The statutes of the IPO determine that the Steering Board decides on the distribution of prizes. Since the time for the evaluation of the essays by the members of the Steering Board is limited the number of medals is limited to a maximum of about 20 essays. In the later years, it has been a practice for the International Jury to make a proposal about which essays should be taken into consideration for a medal and also a proposal about which essays should be awarded an honorable mention.

2. The five criteria of evaluation

The essays are to be evaluated according to the following five criteria of evaluation: relevance to the topic, philosophical understanding of the topic, persuasive power of argumentation, coherence, and originality. Their status is different. And each of them may be interpreted differently.

Relevance. The criterion of relevance is in some sense the most fundamental. If an essay is off topic, then it should not get an award even if it satisfies the other criteria perfectly. For example, if the quotation is about a topic in political philosophy and the essay is only about epistemology, then it is off topic. This does not mean, of course, that an essay about a topic in political philosophy may not also contain epistemological arguments.

The criterion has another, less fundamental application: An essay that is on topic may still contain parts which are less relevant or not relevant to the question the author has chosen to answer. *Ceteris paribus*, the less irrelevant parts the essay contains, the better it is.

Coherence. The criterion of coherence may mean different things. First, it can mean that the essay has a clear and logical structure. Second, it can mean that the terms are clear and used consistently throughout the essay. Third, it can mean that the essay contains claims that are coherent with each other, i.e. which do not contradict each other. The criterion should arguably be considered in all three of these meanings.

Philosophical understanding of the topic: The criterion of philosophical understanding of the topic is allegedly the one about which there are most disagreements among members of the International Jury because it is based on one's understanding of philosophy, and such understanding, as we know, can differ more or less strongly amongst philosophers. Nevertheless, there is consensus among members of the International Jury about some points. First, this criterion applies to the student's understanding of the topic, not necessarily to the topic as it is understood by the author or by the contemporaries of the author. The student may give a correct interpretation of the quotation and this will count as philosophical understanding. But the student may also develop her own thoughts based on the quotation which are not in accordance with what the author of the quotation may have originally meant, and the student may thereby very well show philosophical understanding. Second, knowledge of philosophical claims from the history of philosophy may show philosophical understanding of the topic only if these claims fit into the argumentation of the essay. Superficially reproducing well known philosophical claims or, worse, simple "name dropping" does not, by itself, show philosophical understanding. It is only when the claims are part of an argumentation that they count as philosophical understanding of the topic. Deeper philosophical understanding is shown by the correct and detailed explanation of philosophical claims as well as by the introduction and analysis of concepts relevant to the topic.

Power of argumentation: The criterion of the power of argumentation presupposes that the essay argues for a claim (or for several claims). If the essay does not present a strong philosophical argument it is not worthy of an award unless it has some other striking feature such as a finely worked out original viewpoint or a careful conceptual analysis. Given that the essay argues for a claim, the power of the argumentation may vary quite strongly. The power of argumentation is shown in how well the student develops the arguments and in how good the arguments as such are. Furthermore, it can also be seen in the introduction of possible objections and counter arguments, and in how well these are discussed.

Originality: The criterion of originality is allegedly the most subjective of the five criteria. It means that the essay shows the development of the thoughts of the author and is not simply a repetition of what one can find in textbooks. The criterion of originality does not mean that the essay needs to argue for an unexpected or novel claim.

As has been explicated above, the criteria cannot be applied independently of each other. You are not required to give equal weight to all five of them. When evaluating an essay, it is important to keep in mind the essay as a whole. One useful heuristic method for the general

assessment as well as for remembering the content of the essay is to state its main question and to summarize the answer it gives in one sentence.

3. Discussing essays with other evaluators

It is worth discussing the essay with other evaluators. First, because it allows you to put your own assessment into perspective. Second, because it may help you to see aspects of the essay you may have overlooked. Third, because it is fun to exchange your thoughts with other philosophers.

It is possible that you will disagree on the correct assessment of an essay with another member of the jury. In general, this will be a disagreement among peers as you are both trained philosophers and the arguments of the essays usually do not require detailed knowledge of a particular field of philosophy – although it has to be mentioned that there are exceptions to this, as some previous medal essays have shown. It is itself a philosophical question of how to understand disagreements among peers. There are two opposing views. According to the first one, two parties may disagree and retain their rationality, and according to the second one, any rational disagreement indicates an error on at least one side. This epistemological debate cannot be settled here. As practical advice, I would suggest this: be open-minded, and take the disagreement as an opportunity to learn about the other, yourself, and philosophy!

Acknowledgements: I would like to thank Frank Murphy, Floris Velema, and Erik Gustafsson for very helpful comments about this short guide. It may be revised again in the future. Comments are welcome.

How to cite this article

Pfister, Jonas (2020): Evaluation and Grading of Philosophy Essays at the International Philosophy Olympiad, *Journal of Didactics of Philosophy* 4(3), 165–176.
DOI: 10.46586/JDPh.2020.9588.